2013-11-01

Differential equation models for sharp threshold dynamics

Schramm, Harrison C.

Elsevier Inc.

Mathematical Biosciences 247 (2014) 27-37
http://hdl.handle.net/10945/47535

Mathematical Biosciences 247 (2014) 27–37

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Differential equation models for sharp threshold dynamics

Harrison C. Schramm *, Nedialko B. Dimitrov

Operations Research Department, Naval Postgraduate School, Monterey, CA 93950, United States

ARTICLE INFO

Article history:
Received 7 September 2012
Received in revised form 3 August 2013
Accepted 18 October 2013
Available online 1 November 2013

Keywords:
Differential equation model
Epidemic model
Sharp thresholds

ABSTRACT

We develop an extension to differential equation models of dynamical systems to allow us to analyze probabilistic threshold dynamics that fundamentally and globally change system behavior. We apply our novel modeling approach to two cases of interest: a model of infectious disease modified for malware where a detection event drastically changes dynamics by introducing a new class in competition with the original infection; and the Lanchester model of armed conflict, where the loss of a key capability drastically changes the effectiveness of one of the sides. We derive and demonstrate a step-by-step, repeatable method for applying our novel modeling approach to an arbitrary system, and we compare the resulting differential equations to simulations of the system's random progression. Our work leads to a simple and easily implemented method for analyzing probabilistic threshold dynamics using differential equations.
Published by Elsevier Inc.

1. Introduction

Differential equation models have wide applicability in the study of dynamic systems. They are attractive because they are numerically fast and tractable, transparent in the sense that it is easy to understand how the inputs directly relate to the outputs, and frequently have special cases for which closed-form solutions exist. Our research adapts these models to include systems with stochastic, sharp thresholds. One example of a system with a sharp threshold is a computer network where malicious code is introduced, subject to probabilistic detection and subsequent eradication. In this system, one instant the malicious code is undiscovered, and the following instant it is discovered; discovery defines the sharp threshold and changes system dynamics by allowing a remedy to be applied. For this and many other system, 'half-thresholds', such as "half discovered" are not physically realizable as they do not refer to any realizable state of the system.

A sharp threshold may also be seen in armed conflict where loss of a key capability fundamentally changes the capability of one or both sides.This may be thought of as loss of a key platform or capability. Statements such as "half attrited" do not refer to a realizable state of the system while statements about the probability distribution of a given system surviving are meaningful.

The current method of handling dynamical systems with sharp thresholds is to appeal to simulation of the threshold event by simulating the entire system's random progression. This is useful because it is easily understood, but is expensive, both in terms of computation and time. Often, many simulation repetitions are required to analyze the average behavior of the system and gain useful insights.

It seems that the threshold process and the differential equation model are irreconcilable, chiefly because the threshold event is not divisible in the sense that its expected state is generally not reachable. Our contribution is to overcome this difficulty in a way that has not been previously shown by applying a mean field approximation to the threshold process. By doing so, we create differential equation models that capture the average performance of systems with probabilistic threshold dynamics.

Our approach is novel in that we incorporate the distribution of the threshold time, which may be dependent on the dynamic system state, to create a representation of the average value of the thresholded process. Our method produces a time-trace of the expected state of the system, as well as an explicitly time-dependent, cumulative distribution of the threshold time.

The advantages to be had are numerous. First, by creating a differential equation model, we are able to verify simulation models by comparing them against derived or numerical results. Second, we may use the fast, cheap, differential equation model to scope complex, expensive simulations. Additionally, as a by-product, the model produces the time-dependent cumulative distribution of the threshold time, which prior to modeling may be expressed in terms of the dynamic system state and therefore may not have explicit time dependence. Finally, after developing the theory, we provide two worked examples, along with a step-by-step tutorial on how to apply this method to any thresholded system with a differential equation model.

The organization of the paper is as follows: In Section 2, we review the applicable literature. In Section 3, we derive our novel methodology by developing a mean-field approximation for spreading malware among a computer network, and extract the

* Corresponding author.
E-mail address: harrison.schramm@gmail.com (H.C. Schramm).

0025-5564/$ - see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.mbs.2013.10.009

step-by-step procedure for applying it to other systems. In Section 4, we apply the step-by-step procedure to the Lanchester model of armed conflict. In Section 5, we provide numerical examples comparing the differential equation models to simulations. In Section 5, we also demonstrate that the differential equations from our novel methodology are fundamentally different from differential equations for a non-thresholded system; in other words, no choice of parameters of the non-thresholded differential equations may replicate the behavior of the thresholded differential equations. Finally, in Section 6, we provide some discussion of and directions for future research.

2. Literature review

The general theory and application of differential equation models for physical and social phenomena is a common topic that spans several disciplines, including applied mathematics, biology, and operations research. Many good overviews of the topic exist; for a general text, we recommend *Differential Equation Models* by Braun et al. [8]. For an overview of basic analysis and solution techniques, we recommend O'Neil [34]. Mean-field approximations are frequently used in physics; for an in-depth overview, see the second chapter of Freericks [15]. An overview of approximation methods for probabilistic methods is given by Darling [11].

Mean-field models of epidemics have a long history, and are covered in detail in [10] (see also [2,1]). For a short overview, we recommend the recent tutorial by Dimitrov and Meyers [12]. Fitting models to data is addressed by Mollison [30], and stochastic epidemics are reviewed in detail by Andersson and Britton [2]. Specific system behaviors related to our research, such as time of discovery thresholds, are addressed by Metz et al. [29]. The distribution of the number of infected individuals at the moment of first detection is studied by Trapman and Bootsma [37].

The application of infectious disease models to computer infections has been recommended by Jason [20], an independent group of scientists advising the United States government. A related and noteworthy reference is the case study of the Code Red worm by Moore et al. [31]. Epidemics on networks, and particularly the *S–I–R* model is studied by Draief et al. [13], and the implications of several theoretical network topologies are examined. The spread of malware on wireless networks is considered by Hu et al. [19]. This paper examines the performance of malware—and the steps taken to prevent it—on representative topologies for seven urban ares in the United States.

The work most closely related to our application in malware infections is [38]. Their model closely matches the dynamics of ours in that machines may be in two competing states—infected or patched—and the system operator wishes to maximize the number of patched machines. Our work differs from theirs because, in our model, the detection event occurs as a function of the infection process.

Two recent books by Newman [33,32] describe the formulation and analysis of network models and include cases of epidemics spread on networks as well as the general theory of mean-field approximations. For the specific application of computer infection, our work is different in that we consider both epidemic detection and spread simultaneously in a single, integrated set of differential equations that track the whole progression of the epidemic. Epidemics on networks are also considered in Keeling and Eames [22], where the force of infection changes as a function of time. This work [22] bears some similarity to the Lanchester example we use in this manuscript–in which an equation parameter changes after a certain random threshold event–however, the methods we develop address general changes in dynamics, beyond parameter changes. In Section 3.4, we describe a general, step-by-step method in which system dynamics can change arbitrarily as a function of a threshold event. For example, we could study a system where one set of machines, *I*, infect another set of machines, *S*, then based on some random event that itself depends on the numbers of *I* and *S* machines the dynamics switch and now *S* machines infect *I* machines. This extreme example illustrates the broad applicability and potential of our methodology, in that it can handle whole-sale changes of the entire system, where the timing of the change itself is dependent on the system state.

Mean-field models have been applied in epidemic models of network infections by Lelarge and Bolot [27], and a development of their applicability to general infectious disease models is given in [24], who justify the use of the mean-field approximation for sufficiently large, non-homogeneous networks. Aparicio and Pascual [4] describe the mean-field approach and demonstrate specific cases where it diverges from more sophisticated methods. Essentially, mean field, also called "compartmental" or "mass action" models assume homogeneous mixing. In large networks this may be unsatisfactory because it ignores the dynamic that individuals tend to interact more with those who are close to them than those who are far away. This dynamic is further complicated by considering epidemics on information systems, where the network topology may have little correlation with physical topology. Keeling [21] develops a correction to the basic *S–I–R* model by replacing the stationary infectivity parameter, β with a non-stationary $\beta(t)$, which captures the effect of faster transmission in the beginning stages of infection and slower transmission at the end of the epidemic. Keeling and Eames [22] also addresses the applicability of mean-field models to network epidemic transmission. Though, the focus of this work is a general procedure of creating mean-field models as opposed to the specific application of mean-field models to networks.

Infections with multiple stages are considered by Gani [16]. Their approach partitions the infective population into two subpopulations; a susceptible individual becomes fully part of the infective class after having contact with both the first- stage and second-stage of the infection.

Multiple Stages are also developed by Chowell [9]. These approaches increase the partitioning of the model or generalize the distribution time in each partition but do treat the threshold as global. Finkelstein [14] also studies epidemics with multiple stages. They share our approach of making the first stage of their model the unencumbered, or basic *S–I–R* model. The second stage they consider measures the cost associated with the epidemic. These two stages are analyzed iteratively to determine the amount of prophylactic vaccination that minimizes the epidemics' impact - from a cost perspective. Aggressive prophylactic measures, such as vaccination in biological epidemics are considered by House and Keeling [18]. The problem of determining the correct amount of preventative measure is similar to our problem of patch dissemination in a malware context.

Another perspective on multiple stage epidemics is provided by Klepac and Caswell [25]. Here, feedback exists between the infective process and the 'demographic' process—which consists of immigration and birth. Their approach includes processes that work on different time scales operating in the same model; their overall development measures density-dependent disease transmission. Conversely, our development focuses on the global, uniform change in dynamics created by discovery of malware, or loss of a critical combat enabler. Our threshold may be explicitly dependent on time or implicitly dependent on the other state variables. Furthermore, we extend our model approach beyond epidemics to general dynamic systems, using the Lanchester model of aimed fire as an example. It is possible that for specific scenarios, Kelpac and Caswell's model produces similar results.

For differential equation combat models, we recommend the original paper by Lanchester [26]. A model applying these equations to the Ardennes campaign is in [7]. Combat models with a single, deterministic threshold are studied by Schramm [35].

Our method differs from these examples because we treat it as *global* in the sense that the threshold affects all entities in the model, regardless of their status, by changing the interaction rules. While this could be done in a naive way by specifying the time of transition, and using the final state of the first model as the initial conditions for the second, we choose to do so in a way that respects the stochasticity of the threshold event, delivering the expected value of the modeled process under the stochastic threshold. As we demonstrate in the Appendix, our method captures dynamics that are not present in the unthresholded systems, driving home the point that one could not simply choose the correct parameters of the naive system to satisfactorily achieve the same output.

3. Modeling sharp thresholds

In this section, we describe a basic discrete-time, discrete-state system that models the spread of a cyber infection (henceforth, malware). We use mean-field approximations to derive our novel methodology. Finally, in Section 3.4, we step back from our example to derive a general, step-by-step process for repeating the derivation in other systems.

3.1. Markovian discrete-time dynamics

We begin by considering a model of the spread of malware in a finite population of machines in discrete time. For ease of exposition, we use the term *malware* loosely to describe all malicious code that spreads via inter-machine contact, to include worms, viruses, etc. Similarly, we use the term *infected* to mean that a machine has malware somewhere in its operating system.

We begin with a few basic definitions to facilitate the exposition. There is a fixed population of N machines. At any time, a machine may be in one of the following three states:

Class S: a machine is susceptible, in class S, if it is not currently infected, but may become infected if it interacts with an infected machine.
Class I: a machine is infected, in class I, if it is currently infected and may spread the infection by interaction with a machine of class S.
Class R: a machine is removed, in class R, if it is currently not infected and is immune to infection. A machine may join class R from either class S or I by having a patch installed.

As a preventive measure, a system administrator may specifically design or designate m machines as *sentinels*, which are monitored for infection. Malware may only be detected when it infects a sentinel. After detection, anti-malware measures, which we collectively refer to as *patches*, may be developed and distributed.

Our model has two linked processes–predetection spread and postdetection spread-linked by a detection event. Next, we describe a discrete-time, discrete-state mathematical model of infection progression for each process and the detection event.

By the term discrete-state system, we are referring specifically to the Markov population model of the system. The Markov population model, defined in detail in the following sections, has a state space consisting of vectors $(S_t^P, I_t^P, S_t^D, I_t^D, R_t^D) \in \{0, 1, \ldots, N\}^5$. Initially the last three coordinates of the Markov state are 0, because detection has not yet occurred. The process proceeds in a Markovian fashion until detection occurs, itself a Markovian event, at which point the first two coordinates of the state become 0 and Markovian transitions affect the last three coordinates.

3.1.1. Predetection process

In this section, we describe the discrete-time, discrete-state infection process before detection, which is a standard S–I model of infectious diseases (see [10]). We denote the number of predetection infected machines in round t with I_t^P, and predetection susceptible machines in round t with S_t^P. Spread starts at $t = 0$ with I_0 infected machines and S_0 susceptible machines.

The predetection discrete-time infection process proceeds in rounds. During each round, each machine in class I^P selects a partner machine from the population, uniformly at random, for interaction. If the partner machine is of class I^P, no changes occur. If the partner machine is of class S^P, the partner machine transitions from S^P to I^P with probability β. The number of infected and susceptible machines in round t is random, and the evolution of $\left(S_t^P, I_t^P\right)$ forms a Markov chain. We write the conditional expectation of each coordinate in round $t+1$ in terms of the coordinates in round t as:

$$\mathrm{E}\left[S_{t+1}^P | S_t^P, I_t^P\right] = S_t^P - \frac{\beta S_t^P I_t^P}{N} \tag{1}$$

$$\mathrm{E}\left[I_{t+1}^P | S_t^P, I_t^P\right] = I_t^P + \frac{\beta S_t^P I_t^P}{N}. \tag{2}$$

Eq. (2) states that the expected number infecteds in round $t+1$ is the number of infecteds in round t plus the expected number of newly created infecteds, $I_t^P \cdot S_t^P \cdot \frac{\beta}{N}$. Similar reasoning gives the first equation. The expectation expressions are an approximation, assuming a large population size, N, I small relative to N, so that the likelihood of two infected machines choosing the same susceptible machine is negligible.

3.1.2. Postdetection process

In this section, we describe the infection process after detection has occurred, which is similar to the classic S–I–R model (see [10]). When detection occurs, a *patch* is distributed to all machines in the population: this is a piece of code that, if installed, removes any existing infection and makes the machine (s) resistant to any future infections. Each machine adopts the patch independently with probability μ in each round. We denote postdetection infecteds in round t by I_t^D, postdetection susceptibles in round t by S_t^D, and postdetection removeds in round t by R_t^D.

Postdetection dynamics begin immediately after detection. During the round of detection, t_*, infected members of the population remain infected; i.e., $I_{t_*}^D = I_{t_*}^P$. The malware continues to spread; the expected number of newly created infecteds in round $t+1$ is $\frac{\beta I_t^D S_t^D}{N}$. However, both susceptible machines and infected machines are removed with probability μ.

The random variables $\left(S_t^D, I_t^D, R_t^D\right)$ form a Markov chain in a manner similar to the $\left(S_t^P, I_t^P\right)$ variables. Assuming that detection has occurred in round $t_* \leqslant t$, their expectations in round $t+1$ are:

$$\mathrm{E}\left[S_{t+1}^D | S_t^D I_t^D R_t^D\right] = S_t^D - \frac{\beta S_t^D I_t^D}{N} - \mu S_t^D$$
$$\mathrm{E}\left[I_{t+1}^D | S_t^D I_t^D R_t^D\right] = I_t^D + \frac{\beta S_t^D I_t^D}{N} - \mu I_t^D$$
$$\mathrm{E}\left[R_{t+1}^D | S_t^D I_t^D R_t^D\right] = R_t^D + \mu\left(I_t^D + S_t^D\right).$$

The above equations are nearly identical to the classic S-I-R model, except that they include patch installation, a transition from S to R. Like the Lanchester equations discussed in the next Section, the above expectations are not exact. For example, it is possible for

the first equation to produce negative values when μ and β are close to one. This is because the equation is based on a simplification and does not account for the fact that a machine can only do one of apply the patch or be infected in a single round. None the less, this is a standard simplification in mean-field models, for example it is used for natural birth and death in *S–I–R* models [23]. Substituting the correct expression for the expectation on the right hand side would be more accurate, but result in more complex formulas in the subsequent differential equations. The variables never go negative when we transition to continuous time because β and μ are then interpreted as rates.

3.1.3. The detection event

We are now ready to consider the detection event probabilistically, which is our main contribution. During each round, the m sentinels have the opportunity to contract and detect the malware. We assume that the m sentinels are reselected in a uniform random manner from the population of N machines in each round; this simplifies the model by removing the necessity to track the number of infected sentinels. Detection at an infected sentinel occurs probabilistically. Let α be the probability that a single infected sentinel *does not detect* the malware in a single time period. This choice of parameterization will prove useful in the following development. Let D_t be an indicator random variable of detection:

$$D_t = \begin{cases} 1 & \text{if detection has occurred by time } t \\ 0 & \text{otherwise.} \end{cases}$$

Consider the sequence of differences, $D_t - D_{t-1}$. Members of this sequence are equal to zero everywhere, except in the round of detection. For the round of detection, when t equals t_*, the difference is equal to one. We write the expectation of this difference as

$$\begin{aligned} \mathrm{E}[D_t - D_{t-1}] &= \Pr[D_{t-1} = 0] \cdot \mathrm{E}[D_t - D_{t-1} | D_{t-1} = 0] + \Pr[D_{t-1} = 1] \\ &\quad \cdot \mathrm{E}[D_t - D_{t-1} | D_{t-1} = 1]. \end{aligned} \tag{3}$$

The second term in Eq. (3) is equal to zero because if detection occurred by round $t-1$, it has also occurred by round t. Because the difference expression is nonzero only if detection occurs in round t, we can rewrite the first term in the expression as

$$\Pr[D_{t-1} = 0] \cdot \mathrm{E}[D_t - D_{t-1} | D_{t-1} = 0] = \Pr[D_t = 1, D_{t-1} = 0]. \tag{4}$$

The right-hand side of Eq. (4) expresses the probability that detection occurs on round t and has not occurred in rounds 1 through $t-1$. Given I_i^P, the probability that detection occurs in round i is $1 - \alpha^{I_i^P m/N}$, where the exponent comes from computing the expected number of infected sentinel machines if sentinel machines are chosen uniformly from the population. Let $I_{t:0}^P$ denote the sequence $I_t^P, \ldots, I_0^P$. From Eqs. (3) and (4), we have

$$\mathrm{E}[D_{t+1} - D_t | I_{t:0}^P] = \left(1 - \alpha^{I_t^P m/N}\right) \prod_{k=0}^{t-1} \alpha^{I_k^P m/N}, \tag{5}$$

which is the fundamental difference equation for the D event in this example.

3.2. Coupling the postdetection and predetection processes

The key to properly modeling the sharp change in infection dynamics is the coupling of the predetection process and the postdetection process, as governed by the random detection event. In the round of detection, it is necessary to move all machines from the predetection dynamics to the new postdetection dynamics. For a graphical depiction of coupling, see Fig. 1.

Let (S_t, I_t) be a Markov chain that evolves as an undetected predetection process for all t. Using the notation in the previous sections for predetection and postdetection machines, we can then write the coupled predetection process as

$$\begin{aligned} S_t^P &= (1 - D_t) S_t \\ I_t^P &= (1 - D_t) I_t. \end{aligned}$$

Intuitively, the above expressions say that before detection has occurred—when D_t is 0—the predetection process evolves as specified in Section 3.1.1; and after detection has occurred—when D_t is 1—there are no machines in the S^P and I^P classes. For brevity, let

$$\mathbf{A_t} = (S_t, I_t, S_t^P, I_t^P, D_t, D_{t-1}, S_t^D, I_t^D, R_t^D)$$

denote the state of a Markov chain describing evolution of the cyber infection. We can write the evolution of the coupled postdetection process as

$$\begin{aligned} \mathrm{E}\left[S_{t+1}^D | \mathbf{A_t}\right] &= (D_t - D_{t-1}) S_t + S_t^D - \frac{\beta S_t^D I_t^D}{N} - \mu S_t^D \\ \mathrm{E}\left[I_{t+1}^D | \mathbf{A_t}\right] &= (D_t - D_{t-1}) I_t + I_t^D + \frac{\beta S_t^D I_t^D}{N} - \mu I_t^D \\ \mathrm{E}\left[R_{t+1}^D | \mathbf{A_t}\right] &= R_t^D + \mu\left(I_t^D + S_t^D\right). \end{aligned}$$

Intuitively, these say that in the round when detection occurs—the only time that $D_t - D_{t-1}$ is 1— a sudden inflow of of machines, equal to the machines in the undetected I and S classes, comes into the postdetection classes. Afterward, the postdetection classes behave as described in Section 3.1.2. This allows us to write the complete discrete time difference equations,

$$\mathrm{E}[S_{t+1} | \mathbf{A_t}] - S_t = -\frac{\beta S_t I_t}{N} \tag{6a}$$

$$\mathrm{E}[I_{t+1} | \mathbf{A_t}] - I_t = \frac{\beta S_t I_t}{N} \tag{6b}$$

$$\mathrm{E}[D_{t+1} | I_{t:0}^P] - \mathrm{E}[D_t | I_{t:0}^P] = \left(1 - \alpha^{\frac{I_t^P m}{N}}\right) \prod_{k=0}^{t-1} \alpha^{\frac{I_k^P m}{N}} \tag{6c}$$

$$S_t^P = (1 - D_t) S_t \tag{6d}$$

$$I_t^P = (1 - D_t) I_t \tag{6e}$$

$$\mathrm{E}\left[S_{t+1}^D | \mathbf{A_t}\right] - S_t^D = (D_t - D_{t-1}) S_t - \frac{\beta S_t^D I_t^D}{N} - \mu S_t^D \tag{6f}$$

$$\mathrm{E}\left[I_{t+1}^D | \mathbf{A_t}\right] - I_t^D = (D_t - D_{t-1}) I_t + \frac{\beta S_t^D I_t^D}{N} - \mu I_t^D \tag{6g}$$

$$\mathrm{E}\left[R_{t+1}^D | \mathbf{A_t}\right] - R_t^D = \mu\left(I_t^D + S_t^D\right). \tag{6h}$$

The superscript of D is I^P instead of I because the difference $\mathrm{E}[D_{t+1} | I_{t:0}^P] - \mathrm{E}[D_t | I_{t:0}^P]$ is zero after detection occurs. As written, after detection, I^P is zero, and the difference $\mathrm{E}[D_{t+1} | I_{t:0}^P] - \mathrm{E}[D_t | I_{t:0}^P]$ is zero. If I were used, the difference would never be zero because the I process is not affected by detection, and asymptotically approaches the total population as t increases.

3.3. Moving to continuous time

To move to continuous time from the unit time, discrete difference equations, (6a)–(6h), we create a sequence of random processes, each evaluated at an arbitrarily small time interval, Δt. In each of these processes, we scale the parameters β, μ, and α so as to keep the expected number of events per unit time constant; in continuous time μ and β are properly interpreted as rates and are no longer restricted to the interval $[0, 1]$.

The parameter β gives the expected number of new infections per infected machine per unit time if all other machines are susceptible. For a process that proceeds in time intervals of Δt, the parameter should be scaled to $\Delta t \beta$ because the faster-moving process has $\frac{1}{\Delta t}$ attempts at infection per unit time. Similar reasoning shows that the parameter μ should be scaled to $\Delta t \mu$.

The correct scaling for the parameter α is more delicate. In the unit time progression process, a single infected sentinel does not

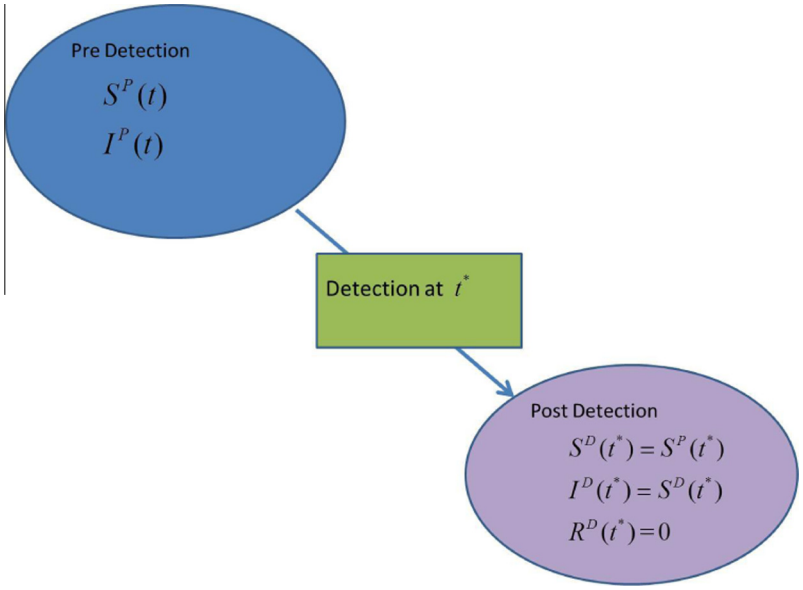

Fig. 1. Depiction of Coupling. In the round of detection, which we call t^*, the state of the predetection process is transferred to the postdetection process.

detect the infection with probability α in each round. The scaling of the parameter should be such that the probability an infected sentinel does not detect the infection remains α for a unit of time. Let α_Δ denote the scaled parameter. The property we seek is $\alpha_\Delta^{\frac{1}{\Delta t}} = \alpha$, because, an infected sentinel has $\frac{1}{\Delta t}$ attempts at detection during a unit time step. Preserving the desired property gives us a scaling of $\alpha_\Delta = e^{\Delta t \ln(\alpha)}$.

The next step of the derivation involves the mean-field assumption, which equates a random variable to its expectation. In general, this step is controversial because it is a heuristic argument, in the sense that it is not explicitly predicated on taking limits of random processes using tools like the functional central limit theorem [6,3]. On the other hand, it is possible to rigorously derive convergence results based on these heuristic approaches, at the cost of a significant increase in mathematical complexity [28]. It is even possible to derive results on the variance of the stochastic process from the means described by the differential equations [5]. Practically, many researchers jump directly to the differential equation models, without considering the underlying Markov chain at all [23,32]. We are explicit in the heuristic limiting argument, without predication on functional central limit theorem, and numerically check the accuracy of the resulting differential equation against a simulation of the Markov chain in Section 5.

With appropriately scaled parameters, we begin with the discrete time difference equations for a process moving in time steps of Δt, apply the mean-field assumption equating a random variable to its expectation, and take the limit as Δt approaches zero to derive the continuous time differential equations. We can follow these steps for each of the Equations (6a)–(6 h), but for exposition we give a few examples highlighting the important details.

For Eq. (6a), the process moving at Δt time intervals has the equation

$$\mathrm{E}[S_{t+\Delta t}|\mathbf{A_t}] - S_t = -\frac{\Delta t \beta S_t I_t}{N}.$$

Applying the mean-field assumption, and dividing both sides by Δt,

$$\frac{S_{t+\Delta t} - S_t}{\Delta t} = -\frac{\beta S_t I_t}{N}.$$

Taking the limit of both sides as Δt goes to zero, we have

$$\frac{dS(t)}{dt} = -\frac{\beta S(t) I(t)}{N},$$

which is the final continuous time equation for the S class. The equation for the I class can be derived similarly.

Deriving the continuous time equation for D is slightly more involved, but consists of the same set of steps. First, we begin with the difference equation for a process moving at Δt time intervals,

$$\mathrm{E}[D_{t+\Delta t}|I^P_{t:0}] - \mathrm{E}[D_t|I^P_{t:0}] = \left(1 - e^{\Delta t \ln(\alpha) I_t^P m/N}\right) \prod_{k=0}^{t-1} e^{\Delta t \ln(\alpha) I_k^P m/N}.$$

Applying mean-field, dividing both sides by Δt, and taking the limit as Δt approaches zero, we have

$$\lim_{\Delta t \to 0} \frac{D_{t+\Delta t} - D_t}{\Delta t} = \lim_{\Delta t \to 0} \frac{\left(1 - e^{\Delta t \ln(\alpha) I_t^P m/N}\right) e^{\Delta t \ln(\alpha) \frac{m}{N} \sum_{k=0}^{t-1} I_k^P}}{\Delta t}.$$

We apply L'Hopital's rule on the right-hand side to derive

$$\frac{dD(t)}{dt} = \lim_{\Delta t \to 0} \left[-\ln(\alpha) \frac{I_t^P m}{N} e^{\Delta t \ln(\alpha) \frac{m}{N} \sum_{k=0}^{t} I_k^P} + \left(1 - e^{\Delta t \ln(\alpha) I_t^P m/N}\right) e^{\Delta t \ln(\alpha) \frac{m}{N} \sum_{k=0}^{t-1} I_k^P} \left(\ln(\alpha) \frac{m}{N} \sum_{k=0}^{t-1} I_k^P\right) \right] = -\ln(\alpha) \frac{I_t^P m}{N},$$

which gives the final continuous time equation for the D variable.

To finish deriving the continuous time system, Eqs. (6d) and (6e) directly translate into their continuous time equivalents.

$$S^P(t) = (1 - D(t)) S(t),$$
$$I^P(t) = (1 - D(t)) I(t).$$

However, for the purposes of a uniform presentation, it may be desirable to represent these as differential equations, which, by the chain rule are:

$$\frac{dS^P(t)}{dt} = -\frac{dD(t)}{dt} S(t) + (1 - D(t)) \frac{dS(t)}{dt},$$
$$\frac{dI^P(t)}{dt} = -\frac{dD(t)}{dt} I(t) + (1 - D(t)) \frac{dI(t)}{dt}.$$

Finally, Eqs. (6f)–(6h) can be converted into continuous equivalents through the standard route of applying the mean-field approximation and taking limits to derive the complete continuous system of equations for the sharp threshold process:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \tag{7a}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N}, \tag{7b}$$

$$\frac{dD}{dt} = -\ln(\alpha)\frac{I^P m}{N}, \tag{7c}$$

$$\frac{dS^P}{dt} = -\frac{dD}{dt}S + (1-D)\frac{dS}{dt}, \tag{7d}$$

$$\frac{dI^P}{dt} = -\frac{dD}{dt}I + (1-D)\frac{dI}{dt}, \tag{7e}$$

$$\frac{dS^D}{dt} = \frac{dD}{dt}S - \frac{\beta S^D I^D}{N} - \mu S^D, \tag{7f}$$

$$\frac{dI^D}{dt} = \frac{dD}{dt}I + \frac{\beta S^D I^D}{N} - \mu I^D, \tag{7g}$$

$$\frac{dR^D}{dt} = \mu\left(I^D + S^D\right), \tag{7h}$$

where we have dropped the explicit dependence on t for brevity. The initial conditions for these equations place the starting number of infected machines in $I(0)$ and $I^P(0)$, the starting number of susceptible machines in $S(0)$ and $S^P(0)$, and set the start of all other variables to zero. By definition $0 \leqslant \alpha \leqslant 1$; the RHS of Eq. (7c) will always be positive.

3.4. Discussion

To gain some understanding of Eqs. (7a)–(7h), we discuss their intuitive interpretation and the general steps to derive them for different sharp-threshold random processes.

Eqs. (7a) and (7b) are tracking a prethreshold process as though the threshold does not exist. Eq. (7c) and the variable D provide a probability distribution (pdf) for the threshold time and $D(t)$ is the cumulative distribution of threshold time. The value of $\frac{dD}{dt}$ can be interpreted simply as the probability density function of the random threshold. For this particular example, $D(t)$ possesses a closed-form solution (see Appendix B).

Eqs. (7d) and (7e) capture the expected trajectories that remain prethreshold. This can be seen in two ways, first by considering the equations $S^P = (1-D)S$ and $I^P = (1-D)I$. The factor $(1-D)$ represents the probability that threshold has not occurred, and only those trajectories where threshold has not occurred stay in the prethreshold classes. The corresponding derivatives in Eqs. (7d) and (7e) also have natural interpretations. The first term subtracts any trajectories where threshold instantaneously occurs. The second term dampens the rate of change, making sure it is proportional to the trajectories where threshold has not occurred.

Eqs. (7f)–(7h) captures the trajectories post-threshold. The first term represents the instantaneous inflow of new trajectories, while the second term computes the change for the postthreshold dynamics. There is no direct inflow of prethreshold trajectories into the R^D class. Similarly, there is no need to dampen the postthreshold changes, the second and third terms of (7f) and (7g), as they are naturally dampened by the fact that only the trajectories that inflow postthreshold are used to compute postthreshold changes. The general steps to derive similar sharp threshold equations for other systems are:

1. Write an unencumbered prethreshold system of equations. In our example, this is Eqs. (7a) and (7b), and variables S and I.
2. Define a variable D to describe the cumulative distribution function of threshold time. Its differential with respect to time is the probability density function for the threshold time, in our example Eq. (7c) Depending on the application, it may be easier to work with D or dD/dt. This function may depend explicitly on t or, as in our example, the expected prethreshold variables, S^P and I^P.
3. Set the expected prethreshold variables to be $(1-D)$ times the unencumbered prethreshold variables. This also defines differential equations of the expected prethreshold variables with respect to t. This is the equivalent of Eqs. (7d) and (7e).
4. Write a postthreshold system of equations. Add terms of $\frac{dD}{dt}$ times the unencumbered variables for direct inflow due to threshold occurrence. This is the equivalent of Eqs. (7f)–(7h).

As we demonstrate computationally in Appendix A, these steps are necessary to correctly track sharp threshold dynamics. Without a similar approach, as adding the unencumbered system and the D variable, there is insufficient state memory to capture sharp threshold dynamics, and we see inaccuracies in the deterministic predictions versus the expected state of the underlying random process.

4. Application to Lanchester equations

In this Section, we employ steps 1–4 from the previous section to develop a novel thresholded model of combat based on Lanchester's equations. This model is useful both on its own as a contribution to combat models, as well as an example of steps 1–4 in Section 3.4.

Our mathematical model considers an immediate, global loss of effectiveness for one of the combatants. Such a loss of effectiveness may come from the loss of a key enabler, such as communication networks, or vital asset, and could be the result of mechanical failure or adversary action.

Lanchester [26] involves two opposing forces, which we call blue and red. The total amount of blue forces available at time t is denoted by the variable $B(t)$, and the total amount of red forces available at time t is denoted by the variable $R(t)$. Lanchester's *aimed fire* equations assume that each red unit has a likelihood of ρ of removing a blue unit, while each blue unit has a likelyhood β of removing a red unit. Lanchester describes the evolution of the battle as:

$$\frac{dB(t)}{dt} = -\rho R(t), \tag{8a}$$

$$\frac{dR(t)}{dt} = -\beta B(t), \tag{8b}$$

where ρ, β are effectiveness parameters of the red (blue) sides, respectively. These equations have been well studied and applied to numerous case studies (see [39]). We generalize this model to consider cases where one of the effectiveness parameters, say β is suddenly and irrevocably reduced its prethreshold value to a lower, postthreshold value, say β^-.

To create the threshold model of the Lanchester equations, we follow steps 1–4 outlined in Section 3.4.

1. We write the unencumbered prethreshold system of equations. In this case, they are identical to Lanchester's original formulation, Eqs. (8b).

2. We define a variable $D(t)$, that describes the cumulative distribution function of threshold time. In this example, we choose an exponentially distributed threshold time with rate parameter λ. This gives $\frac{dD(t)}{dt} = \lambda e^{-\lambda t}$, and $D(t) = 1 - e^{-\lambda t}$, with $D(0) = 0$.
3. We write the prethreshold equations, dropping the dependence on t for brevity,

$$\frac{dB^P}{dt} = -\frac{dD}{dt}B + (1-D)\frac{dB}{dt},$$
$$\frac{dR^P}{dt} = -\frac{dD}{dt}R + (1-D)\frac{dR}{dt}.$$

As in our previous infection example, these equations result from setting $B^P = (1-D)B$ and differentiating.

4. We now write the postthreshold equations, modeling inflow by adding terms of $\frac{dD}{dt}$ where appropriate:

$$\frac{dB^D}{dt} = \frac{dD}{dt}B - \rho R^D,$$
$$\frac{dR^D}{dt} = \frac{dD}{dt}R - \beta^- B^D.$$

The four steps generate the complete set of differential equations

$$\frac{dB}{dt} = -\rho R, \tag{9a}$$
$$\frac{dR}{dt} = -\beta B, \tag{9b}$$
$$\frac{dD}{dt} = \lambda e^{-\lambda t}, \tag{9c}$$
$$\frac{dB^P}{dt} = -\frac{dD}{dt}B + (1-D)\frac{dB}{dt}, \tag{9d}$$
$$\frac{dR^P}{dt} = -\frac{dD}{dt}R + (1-D)\frac{dR}{dt}, \tag{9e}$$
$$\frac{dB^D}{dt} = \frac{dD}{dt}B - \rho R^D, \tag{9f}$$
$$\frac{dR^D}{dt} = \frac{dD}{dt}R - \beta^- B^D. \tag{9g}$$

The required initial conditions are: $B(0)$ and $B^P(0)$ as initial blue forces, $R(0)$ and $R^P(0)$ as initial red forces, and all other variables to zero.

5. Numerical analysis

In this section, we compare our theoretical results with simulations to verify that the differential equations do indeed track the average state of the underlying Markov chain. This is a critical step in verifying the differential equation models because mean-field approximations assume equality between a random variable and its mean—and thus provide no mathematical guarantee on the result. We do this numerical analysis and verification for both the malware model developed in Section 3 and the Lanchester model of Section 4. We also demonstrate that the models we develop are fundamentally different than the original systems of differential equations by showing that no parameterization of the original differential equations yields correct behavior.

Fig. 2 depicts a comparison of a simulation of cyber infection to thresholded model as presented in Eqs. (7). Both the simulation and the differential equations use a $(\beta, \alpha, m, N, \mu) = (0.01; 0.99; 20; 100,000; 0.2)$ and 100 initially infected machines. The dashed lines indicate the average state of 2000 simulation runs; i.e., the average state of the Markov chain at time t, for each of the predetection and postdetection classes. The solid lines with markers indicate the numerical integration of the differential equations. For all pre and postdetection classes, the average of the simulation runs agrees with the differential equations. This choice of parameterization, in particular $\mu = 0.2$, results in highly variable postdetection classes, S^D and I^D. This variance is a result of the quick adoption of the patch after detection has occurred. The plots of S^D and I^D indicate a benefit of the differential equations—that they can produce the mean state of the system without the requirement for thousands of simulations. Fig. 2 depicts agreement between the empirical distribution of detection time, derived from the simulation and pictured as a histogram, versus the variable D in the differential equation system. This demonstrates another benefit to the differential equation system: the differential equation system

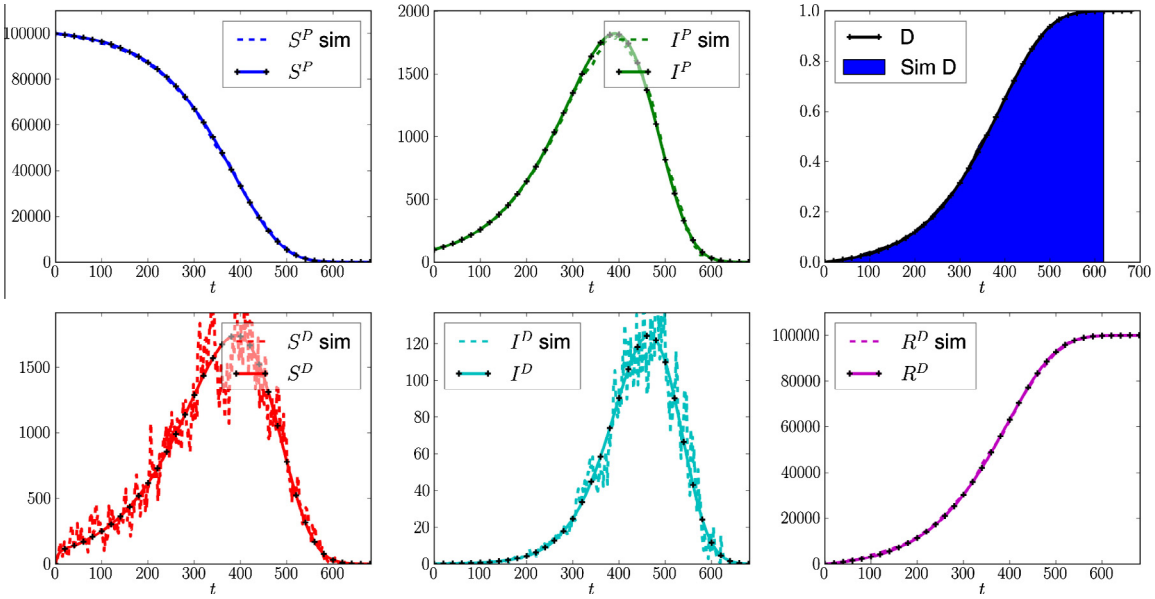

Fig. 2. Cyber infection model simulation versus differential equations. Both methods use a parameterization of $(\beta, \alpha, m, N, \mu) = (0.01; 0.99; 20; 100,000; 0.2)$ and 100 initially infected machines. The dashed lines indicate the average state of 2000 simulation runs, and the solid lines with markers indicate the numerical integration of the differential equations. The parameter $\mu = 0.2$ models quick adoption of the patch and results in variable postdetection classes, S^D and I^D. The differential equations produce the mean state of the system accurately, even with this variance. The top right graph depicts agreement between the empirical distribution of detection time, derived from the simulation and pictured as a histogram, versus the variable D in the differential equation system. In this process, the detection time is a function of the system state.

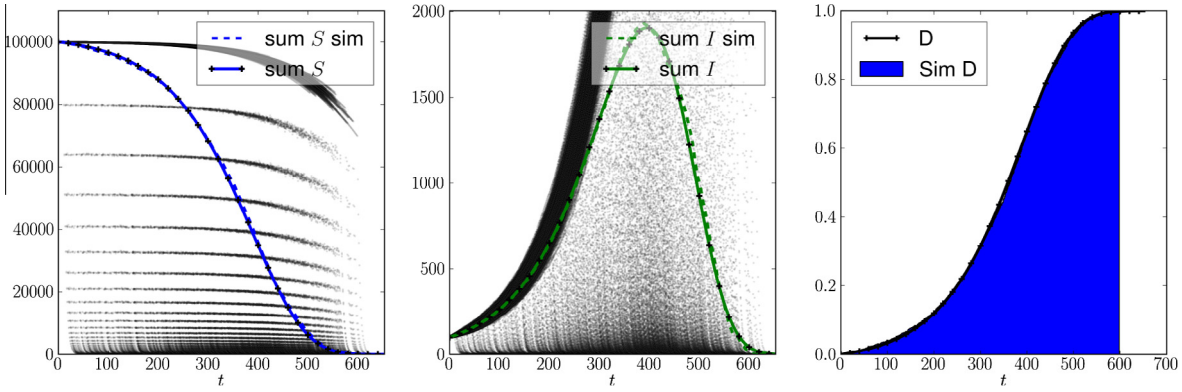

Fig. 3. Total susceptibles and infecteds in cyber infection model. The dashed lines indicate the average state of 2000 simulation runs, and the solid marked lines indicate the result of the differential equations, $S^P + S^D$ and $I^P + I^D$. The black dots depict a scatter plot of the state of the 2000 simulations. The models use the same parameterization as in Fig. 2. The striations in the scatter plot for susceptibles is due to the fast adoption of the patch, $\mu = 0.2$. Approximately 20% of machines adopt the patch in each round. The differential equation system accurately captures the average state of the system.

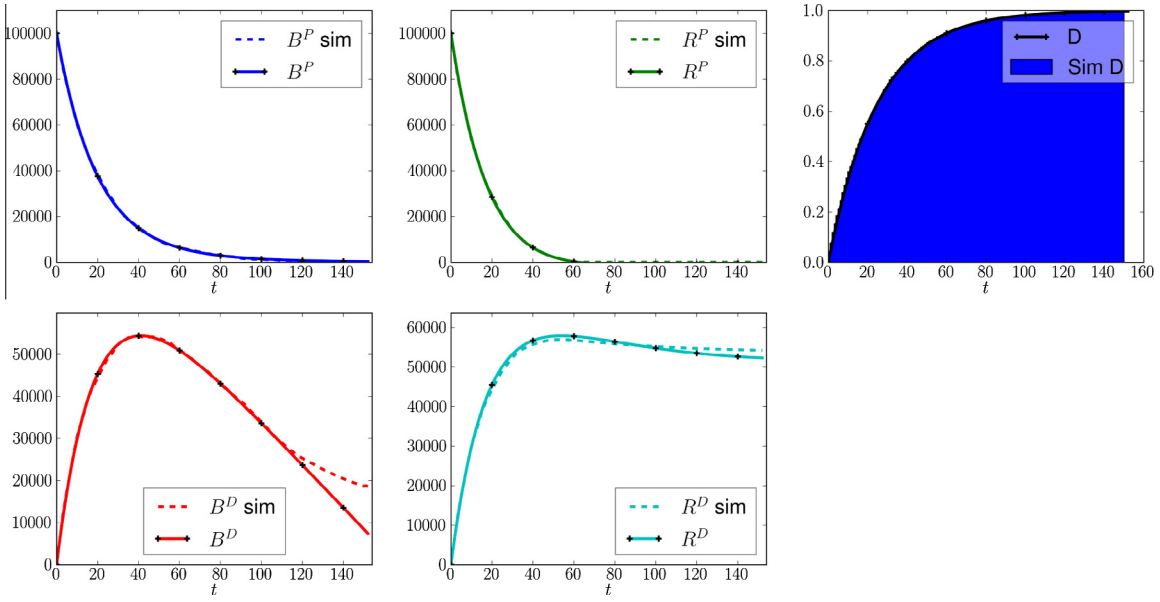

Fig. 4. Lanchester combat model simulation versus differential equations. Both methods use a parameterization $(\rho, \beta, \beta^-, \lambda) = (0.01, 0.02, 0.001, \frac{1}{25})$ and initial sizes of 100,000 for both forces. The standard Lanchester model has inaccuracies at high values of t that give the disagreement between the simulations and the differential equations for the class B^D at high values of t. For small values t, less than approximately 120 in the figure, the average state of the simulations agrees with the differential equations.

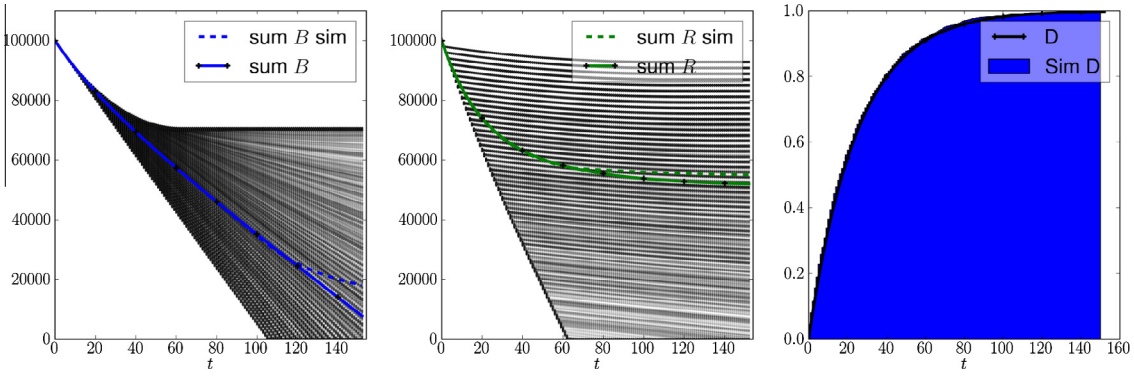

Fig. 5. Total force sizes in Lanchester combat model. The dashed lines depict the average of 2000 simulations, while the solid marked lines depict the sums $B^P + B^D$ and $R^P + R^D$. The scatter plot depicts states of the 2000 simulation runs. The models use the same parameterization as Fig. 4. For an individual simulation, the red forces follow the sharp down curve, until threshold time, at which point they follow one of the flatter striations. Even with the highly variable force sizes between individual simulations, Eq. (9) accurately captures the expected force sizes at time t.

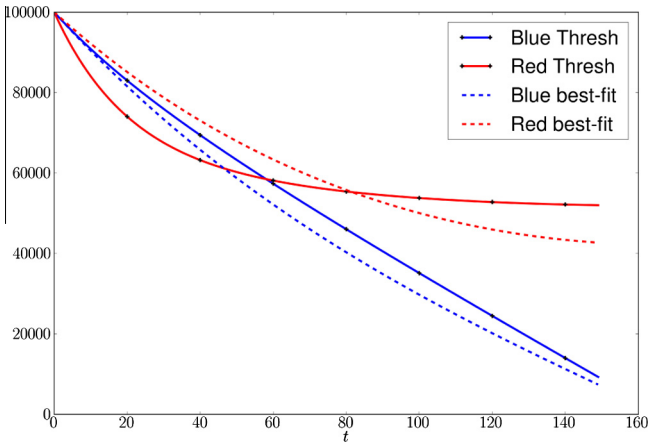

Fig. 6. Best fit of a standard Lanchester model to a sharp threshold Lanchester model. The solid lines represent the expected size of the red and blue forces under Eq. (9), while the dashed lines represent the closest fitting parameterization of Eq. (8), standard Lanchester model. The best fit optimization finds the parameter β for the standard Lanchester model that minimizes the squared error between the model's force sizes and Eq. (9)'s force sizes. All other parameters for both models are the same as those in Fig. 4. The fit between the closest Lanchester model and our modeling method is poor, demonstrating that the sharp threshold models yield fundamentally new behavior. For this example, the red forces initially diminish rapidly, but after the sharp threshold, overwhelm the blue forces.

can produce a distribution of threshold time even when the threshold time is a function of the system state, as is the case for the cyber infection model.

The model described by Eqs. (7) split pre- and post-detection susceptibles and infecteds into different classes; however, an analyst may be interested simply in the number of susceptibles and infecteds at time t. Fig. 3 depicts a comparison between simulation and differential equations on the number of susceptibles at time $t, S^P + S^D$, and the number of infecteds at time $t, I^P + I^D$. The dashed lines indicate the average state of 2000 simulation runs, and the solid marked lines indicate the result of the differential equations. In addition, the figure includes a scatter plot for the 2000 runs. The variance due to the fast adoption of the patch, $\mu = 0.2$, is evident in the bands in the scatter plot for susceptibles. Once detection occurs, approximately 20% of machines adopt the patch in each round, resulting in the striation in the figure. Even with this large amount of variance in the individual simulation runs, the differential equation system accurately captures the average state of the system.

Fig. 4 depicts a comparison of a Lanchester combat model simulation to the corresponding differential equation model as presented in Eqs. (9g). Both the simulations and the differential equations use a parameterization $(\rho, \beta, \beta^-, \lambda) = (0.01, 0.02, 0.001, \frac{1}{25})$ and initial sizes of 100,000 for both the blue and red forces. The differential equations agree with the mean state of the system, except at higher values of t. This disagreement is due to the well-known inaccuracy of the standard Lanchester aimed fire model as presented in Eqs. (8b) (see [39,36]). The standard Lanchester model is inaccurate because it overestimates the effectiveness of a large force against a small force, possibly even resulting in negative force sizes. The crux of the matter is that the mean-field approximation in the Lanchester model fails when the numbers become small on either side. For large numbers of a force, say X, we may say that $E[X|X > 0] \approx E[X]$. As X becomes small, this assumption begins to fail. A solution to this difficulty is presented by Hartley [17], where the generalized Lanchester equations have an exponent in each variable, e.g., $dR/dt = -\beta B^C R^F$, and our aimed fire example is a special case where $C = 1$, $F = 0$. Intuitively this model works because conflict is rarely exclusively aimed or area.

Fig. 5 depicts the expected size of the blue and red forces at time t. The dashed lines depict the average of 2000 simulations, while the solid marked lines depict the sums $B^P + B^D$ and $R^P + R^D$. The figure also includes a scatter plot of the states of the 2000 simulation runs. The striations in the scatter plot for the red forces is due to variance in the threshold time. Before the threshold, the blue forces are highly effective against red, and after the threshold they become ineffective. For an individual simulation, the red forces would follow the sharp down curve, until threshold time, at which point they would follow one of the flatter striations. Even with the highly variable force sizes between individual simulations, Eqs. (9g) accurately captures the expected force sizes at time t.

Fig. 6 demonstrates that our modeling method is fundamentally different than a simple application of previously existing models. Specifically, consider the Lanchester threshold system, where the sharp threshold simply reduces the effectiveness of the blue forces from β to β^-. One modeling approach may be to simply replace the parameter β in the standard Lanchester model, as presented in Eq. (8b), with an expected effectiveness parameter of the blue forces. In Fig. 6, the solid lines represent the expected size of the red and blue forces under the model presented in Eqs. (9g), while the dashed lines represent the closest fitting standard Lanchester model (see Eqs. (8b)). The best fit standard Lanchester model results from a least-squares optimization on the parameter β, minimizing the difference from the force sizes given by the model in Eqs. (9g). The fit between the closest Lanchester model and our modeling method is poor,, implying that our approach yields fundamentally different behavior than the original system. Appendix A demonstrates that a naive approach—one which does not include the unencumbered system required by Step 1 in Section 3.4—to modeling the more complex problem of cyber infections also does not work.

6. Conclusions and future research

We extend the utility of differential equation models by incorporating the novel ability to model a probabilistic sharp threshold in system dynamics. We demonstrate our results with two applications: modeling cyber infections and capability loss in combat—both of which are of interest in their own right. Our cyber infection model may be useful in understanding the relative merits of investment in additional detectors versus more rapid patch dissemination. Similarly, we hope that our Lanchester extension will be useful in quantifying the tradeoff between protection and attack for critical capabilities. Beyond these two applications, we develop a simple, step-by-step procedure to model sharp thresholds in other systems. The steps described in Section 3.4 provide intuition and allow other modelers to create probabilistic sharp threshold models, without re-creating the steps in Section 3.

Future areas of study that would build on this work include: (1) a broader set of problems against which to apply our novel modeling method; (2) to more fully develop the multiple threshold problem, and (3) to use the differential equations to describe the variance in the underlying Markov chain; the large amount of variance is visible in the numerical analysis for both examples we consider, and it would be interesting and relevant to describe that variance by perhaps using stochastic diffusion approaches. There has been some work on describing the variance around mean-field models [5]. Incorporating these variance tracking techniques into the threshold models we present may be possible, but requires significant effort beyond the scope of our contribution. Developing a simple step-by-step process, similar as the one we outline in the paper for sharp thresholds, for creating equations that track the variance of the process would be of great use to practitioners.

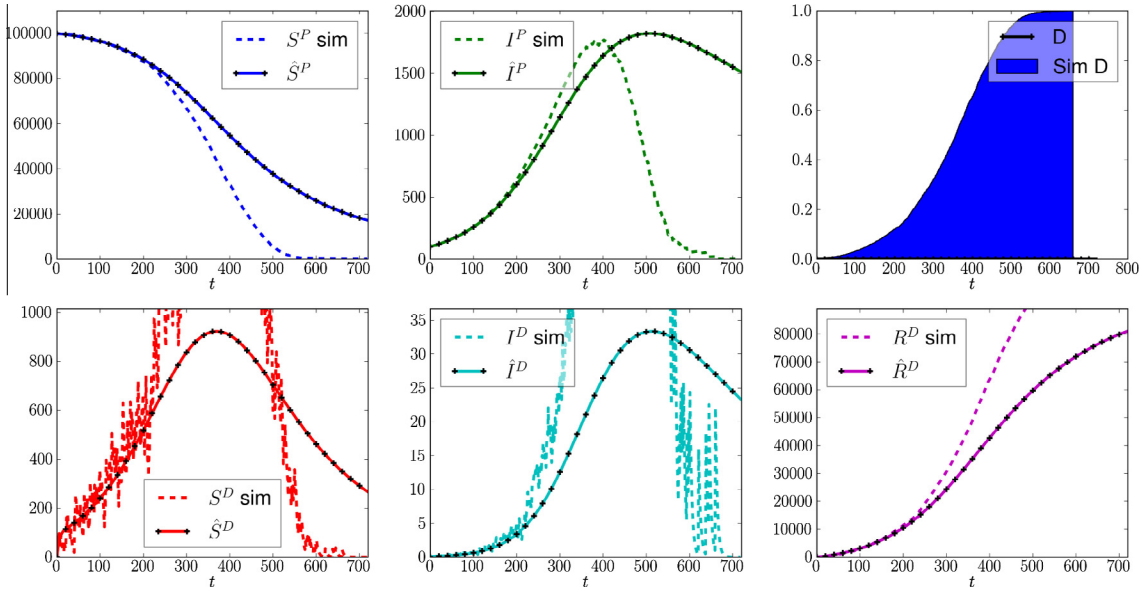

Fig. A.1. Comparison of simulation to naive cyber infection model. These figures parallel those of Fig. 2. The simulation and the naive model, (A.5), are parameterized in the same way as the models in Fig. 2. The naive approach simply does not hold a sufficient amount of state to accurately describe the evolution of the system. The simulation and the naive differential equation model begin in agreement, but quickly drift apart in all state variables.

An additional direction of future work would be to extend the techniques in this paper to address multiple thresholds. A key consideration in multiple thresholds is the joint distribution of the thresholds. In other words, how does the occurrence of one threshold event effect other threshold events? Perhaps the simplest way to start addressing this is to consider independent thresholds, where the occurrence of one does not affect the other. One might attempt to do multiple thresholds by a recursive application of the method described in this paper: apply the method once for the first threshold, then apply the method again on the resulting system for the second threshold. Unfortunately, such a recursive application does not track the two-threshold Markov chain well. Intuitively, the reason that the steps outlined in this paper work are because of the key insight to maintain the unencumbered system. The unencumbered system allows us to track the behavior of all Markov chain trajectories before they experience the sharp threshold. In a two-threshold model, we have not been able to come up with an equivalent way to track the progression of trajectories that experience one threshold but not the other. A new insight is required to move the research forward to multiple thresholds. This new insight should be applicable to multiple thresholds with arbitrary joint distributions.

Appendix A. Naive model for cyber infections

A naive approach to modeling the probabilistic sharp threshold for cyber infections ignores the unencumbered variables described in step 1 in Section 3.4. The naive approach results in the following system of equations:

$$\frac{d\hat{S}^P}{dt} = -\frac{\beta \hat{S}^P \hat{I}^P}{N} + \frac{\ln(\alpha)\hat{S}^P\hat{I}^P m}{N} \tag{A.1}$$

$$\frac{d\hat{I}^P}{dt} = \frac{\beta \hat{S}^P \hat{I}^P}{N} + \frac{\ln(\alpha)\hat{I}^P\hat{I}^P m}{N} \tag{A.2}$$

$$\frac{d\hat{S}^D}{dt} = -\frac{\ln(\alpha)\hat{S}^P\hat{I}^P m}{N} - \frac{\beta \hat{S}^D \hat{I}^D}{N} - \mu \hat{S}^D \tag{A.3}$$

$$\frac{d\hat{I}^D}{dt} = -\frac{\ln(\alpha)\hat{I}^P\hat{I}^P m}{N} + \frac{\beta \hat{S}^D \hat{I}^D}{N} - \mu \hat{I}^D \tag{A.4}$$

$$\frac{d\hat{R}^D}{dt} = \mu\left(\hat{I}^D + \hat{S}^D\right). \tag{A.5}$$

The differential equations of model (A) follow from the difference equations for the underlying Markov chain, and are natural. For example, intuitively, Eq. (A.1) says that the predetection susceptible class decreases either through infection, which occurs instantaneously with probability $\frac{\beta I^P}{N}$, or detection, which occurs instantaneously with probability $-\frac{\ln(\alpha)S^P I^P m}{N}$. The other equations of model (A) can be derived and described similarly.

Fig. A.1 shows that this naive approach does not track the average state of the underlying Markov chain. Both the simulation and model (A) are parameterized with the same parameters as those in Fig. 2. Across all state variables, the differential equation and the simulation begin in agreement, but later drift apart. Intuitively, this is because model (A) reaches states that can never be reached by the simulation. An accurate model requires more state—the unencumbered system that tracks the prethreshold progression—and explicit modeling of the sharp threshold event—the D variable. This intuition leads to the development of the method in Section 3, and results in the correct model presented as Eqs. (7).

Appendix B. Solution of the D equation

Eq. (7c) is equivalent to

$$I^P(t) = (1 - D(t))I(t)$$

It is known that $I(t)$ has a closed solution,

$$I(t) = \frac{I_0 N}{I_0 + S_0 e^{-\beta N t}}.$$

For details see Daley and Gani (1999). We use the equation for $I(T)$ to define $I^P(t)$ in terms of $D(t)$, and substitute the result into (7c) to get

$$\frac{dD}{dt} = \frac{-\ln(\alpha)m}{N}(1 - D)\frac{I_0 N e^{\beta N t}}{I_0 e^{\beta N t} + S_0}.$$

Separating variables gives

$$\frac{dD}{1 - D} = \frac{-\ln(\alpha)m}{N}\frac{I_0 N e^{\beta N t}}{I_0 e^{\beta N t} + S_0}dt,$$

which is valid because $D(t) < 1\ \forall_t$, and therefore $1 - D(t) > 0$. The key step is that both sides of this equation are of the form dU/U. We let $U = 1 - D$ and $V = I_0 e^{\beta N t} + S_0$ to derive

$$\frac{-dU}{U} = \frac{-\ln(\alpha)m}{N\beta}\frac{dV}{V}.$$

Multiplying both sides by -1 and integrating gives

$$\ln(1-D) = \frac{\ln(\alpha)m}{N\beta}\ln(I_0 e^{\beta N t} + S_0) + C.$$

The above expression reduces to

$$D = 1 - \kappa\left[I_0 e^{\beta N t} + S_0\right]^{\frac{\ln(\alpha)m}{N\beta}},$$

where $\kappa = N^{\frac{-\ln(\alpha)m}{N\beta}}$ to ensure the initial condition $D(0) = 0$.

References

[1] R.M. Anderson, R.M. May, Infectious diseases of humans: dynamics and control, Oxford University Press, Oxford; New York, 1991.
[2] H. Andersson, T. Britton, Stochastic epidemic models and their statistical analysis, Springer, New York, 2000.
[3] Hakan Andersson, Boualem Djehiche, A functional limit theorem for the total cost of a multitype standard epidemic, Advances in Applied Probability 26 (3) (1994) 690.
[4] J.P. Aparicio, M. Pascual, Building epidemiological models from r: an implicit treatment of transmission in networks, Proceedings: Biological Sciences 274 (16) (2007) 505–512.
[5] A.D. Barbour, On a functional central limit theorem for markov population processes, Advances in Applied Probability 6 (1974) 21.
[6] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[7] M. Bracken, J. Kress, R.E. Rosenthal, Warfare modeling, Military Operations Research Society, Alexandria, VA, 1995.
[8] M. Braun, C.S. Coleman, D.A. Drew, Differential Equation Models, Springer-Verlag, New York, 1983.
[9] G. Chowell, Mathematical and Statistical Estimation Approaches in Epidemiology, Springer, Dordrecht, 2009.
[10] D.J. Daley, J.M. Gani, Epidemic Modelling: An Introduction, Cambridge University Press, Cambridge, New York, 1999.
[11] R.W.R. Darling, James R. Norris, et al., Differential equation approximations for markov chains, Probability Surveys 5 (2008) 37.
[12] N. Dimitrov, L. Meyers, Mathematical approaches to infectious disease prediction and control, Tutorials in Operations Research, INFORMS, Hanover MD, 2010.
[13] M. Draief, A. Ganesh, L. Massoulié, Thresholds for virus spread on networks, The Annals of Applied Probability 18 (2) (2008) 359–378. ID: 4892180224.
[14] S.N. Finkelstein, A two-stage model for the control of epidemic influenza, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 1980. ID: 8926959.
[15] James K. Freericks, Transport in multilayered nanostructures the dynamical mean-field theory approach, Imperial College Press, London, 2006.
[16] J. Gani, A carrier-borne epidemic with multiple stages of infection, Journal of Applied Probability 28 (1) (1991) 1.
[17] D.S. Hartley, Predicting Combat Effects, INFORMS, Linthicum, MD, 2001.
[18] T. House, M. Keeling, Deterministic epidemic models with explicit household structure, Mathematical Biosciences 213 (1) (2008) 29.
[19] H. Hu, S. Myers, V. Colizza, A. Vespignani, G. Parisi, Wifi networks and malware epidemiology, in: Proceedings of the National Academy of Sciences of the United States of America, vol. 106 (5), 2009, pp. 1318–1323. ID: 4910432646.
[20] JASON and MITRE Corporation, Science of cyber-security, 2010.
[21] M. Keeling, The implications of network structure for epidemic dynamics, YTPBI Theoretical Population Biology 67 (1) (2005) 1.
[22] M.J. Keeling, K. Eames, Networks and epidemic models, Journal of the Royal Society, Interface/ the Royal Society 2 (4) (2005) 295.
[23] M.J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, Princeton, 2008.
[24] A. Kleczkowski, B.T. Grenfell, Mean-field-type equations for spread of epidemics: the 'small world' model, Physica A 274 (1–2) (1999) 355.
[25] H. Klepac, P. Caswell, The stage-structured epidemic: linking disease and demography with a multi-state matrix approach model, Theoretical Ecology 4 (3) (2001) 301–319.
[26] F.W. Lanchester, Aircraft in Warfare, The Dawn of the Fourth Arm, Constable and Company Limited, London, 1916.
[27] M. Lelarge, J. Bolot, A local mean field analysis of security investments in networks, CoRR, abs/0803.3455, 2008.
[28] D.R. McNeil, S. Schach, Central limit analogues for Markov population processes, Journal of the Royal Statistical Society: Series B (Methodological) 35 (1) (1973) 1–23.
[29] J.A.J. Metz, M. Wedel, A.F. Angulo, Discovering an epidemic before it has reached a certain level of prevalence, Biometrics 39 (3) (1983) 765.
[30] D. Mollison, Epidemic Models: Their Structure and Relation to Data, Cambridge University Press, New York, NY, Cambridge England, 1995.
[31] D. Moore, C. Shannon, J. Brown. Code-Red: a case study on the spread and victims of an Internet worm, in: Internet Measurement Workshop (IMW) 2002, pp. 273–284, Marseille, France, Nov 2002. ACM SIGCOMM/USENIX Internet, Measurement Workshop.
[32] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford, New York, 2010.
[33] M.E.J. Newman, A.L. Barabasi, D.J. Watts, The Structure and Dynamics of Networks, Princeton University Press, Princeton, 2006.
[34] P.V. O'Neil, Advanced Engineering Mathematics, Wadsworth Pub. Co., Belmont, CA, 1983.
[35] H.C. Schramm, Lanchester models with discontinuities: an application to networked forces, Military Operations Research 17 (4) (2012) 59.
[36] J.G. Taylor, Lanchester Models of Warfare, Ketron Inc., Arlington, VA, 1983.
[37] P. Trapman, M. Bootsma, A useful relationship between epidemiology and queueing theory: the distribution of the number of infectives at the moment of the first detection, Mathematical Biosciences 219 (1) (2009) 15.
[38] M. Vojnovic, A.J. Ganesh, On the race of worms, alerts, and patches, IEEE/ACM Transactions on Networking 16 (5) (2008) 1066.
[39] A.R. Washburn, M. Kress, Combat modeling, Springer-Verlag, New York, 2009.